

Scanline Sampler without Detailed Balance: An Efficient MCMC for MRF Optimization

Wonsik Kim* and Kyoung Mu Lee

Department of ECE, ASRI, Seoul National University, 151-742, Seoul, Korea

{ultra16, kyoungmu}@snu.ac.kr, <http://cv.snu.ac.kr>

Abstract

Markov chain Monte Carlo (MCMC) is an elegant tool, widely used in variety of areas. In computer vision, it has been used for the inference on the Markov random field model (MRF). However, MCMC less concerned than other deterministic approaches although it converges to global optimal solution in theory. The major obstacle is its slow convergence. To come up with faster sampling method, we investigate two ideas: breaking detailed balance and updating multiple nodes at a time. Although detailed balance is considered to be essential element of MCMC, it actually is not the necessary condition for the convergence. In addition, exploiting the structure of MRF, we introduce a new kernel which updates multiple nodes in a scanline rather than a single node. Those two ideas are integrated in a novel way to develop an efficient method called scanline sampler without detailed balance. In experimental section, we apply our method to the OpenGM2 benchmark of MRF optimization and show the proposed method achieves faster convergence than the conventional approaches.

1. Introduction

Markov random field (MRF) has been used in numerous areas in computer vision [18]. MRFs are generally formulated as follows. Given a graph $G = (\mathcal{V}, \mathcal{E})$, the joint probability function of the pairwise MRF is given by

$$P(\mathbf{x}) \propto \prod_{i \in \mathcal{V}} \phi_i(x_i) \cdot \prod_{(i,j) \in \mathcal{E}} \phi_{i,j}(x_i, x_j), \quad (1)$$

where \mathcal{V} is the set of nodes, \mathcal{E} is the set of edges, and $x_i \in \{1, 2, \dots, L\}$ is the label assigned on node i . Obtaining maximum a posteriori (MAP) solution of probability (1) is NP-hard in general cases. To achieve better approximation solutions, many different optimization methods have been developed.

The Markov chain Monte Carlo (MCMC) is one of the methods used to solve the aforementioned problem. MCMC has been used on many different areas because it is applicable to general posterior models and can find the global optimal solution in theory. The MCMC design has become convenient since the advent of the Metropolis–Hastings (MH) method. MCMC provides a simple scheme to design the kernel that converges to the desired stationary distribution. However, despite its simplicity and theoretical elegance, MCMC is not widely used in the MRF optimization because sampling-based methods are said to be much slower than deterministic-based methods are. In this paper, we investigate two ideas to accelerate MCMC. Section 3 introduces a non-reversible transition kernel that breaks detailed balance. Section 4 presents a method of updating multiple nodes at a time. Finally, Section 5 integrates those two ideas to develop a novel and efficient algorithm called scanline sampler without detailed balance.

2. Background

In 1984, Geman and Geman [7] developed a general Bayesian framework for vision problems where MRF is used to model the prior distribution of the image. Since then, MRF has become one of the most well-formulated models in the entire computer vision field. This model has been used in variety of applications, from low- to high-level vision. In their work, they used Gibbsian simulated annealing (SA) to obtain the MAP solution.

In 2001, Boykov *et al.* [6] proposed the algorithms α -expansion and $\alpha\beta$ -swap, which are considered faster than SA. The proposed algorithms solve multi-label problems in an iterative manner by using graph cuts. They showed that their methods outperformed SA both in terms of speed and quality of the solution.

Many studies have focused on deterministic approaches, such as graph cutbased algorithms [14, 5, 1, 12] and message-passing algorithms [19, 16, 21, 13]. These two algorithms are considered the major approaches in this field. The graph cutbased algorithm was applied to a limited class

*Currently at SAIT, Samsung Electronics

of energy functions. Nevertheless, the solvable class of this algorithm has increased; thus, many existing energy functions are known to be solvable these days.

The message-passing algorithm was applied without a theoretical guarantee to converge at the beginning (*e.g.* loopy belief propagation). By adopting the dual decomposition framework, message-passing algorithms such as tree-reweighted message passing (TRW), can provide the lower bound of the optimal solution. Recently, this lower bound has become increasingly tighter.

Although many studies on deterministic optimization algorithms are available, few studies have focused on MCMC because of its slow convergence. Nevertheless, several attempts have been made to improve the convergence speed of MCMC in MRF optimization context. We will introduce some examples in the following subsection.

2.1. Previous MCMC approaches

One method to increase the convergence speed of MCMC is to sample multiple nodes at a time. On the basis of this aspect, Barbu and Zhu adopted Swendsen–Wang cuts (SWC) [3, 4]. In SWC, a set of nodes with the same label is stochastically chosen at each iteration, and the labels of nodes are stochastically flipped by using MH method. However, SWC is restricted to a specific class of applications because of the inherent characteristics of the SWC kernel. SWC is suitable only for Potts-like energy function, which favors piece-wise constant solutions.

Another method of increasing the convergence speed of MCMC is to adopt population-based framework and parallelize the annealing schedule. To this end, population-based MCMC [11] was proposed. Three different kernels were designed: mutation kernel for the single-chain update, and the crossover and exchange kernels for the interaction between chains. Given that the mutation kernel was designed in the same way as SWC, this approach has the same limitation as that of SWC, but with significant improvement.

MCMC was combined with deterministic algorithms to further improve its performance [7]. In this framework, deterministic methods, such as graph cuts, belief propagation (BP), and TRW, were used for the transition kernel. This approach achieved lower energy solution than other sampling-based and deterministic methods. However, the performance of MCMC depends on the deterministic methods to be combined.

3. Non-reversible kernel (Suwa–Todo method)

Recent studies have shown that MCMC can be accelerated by breaking the detailed balance [17, 20], which has often been considered one of the essential elements of MCMC. Given that detailed balance is not a necessary condition, stationary distribution is achievable even without detailed balance. However, designing MCMC kernel without

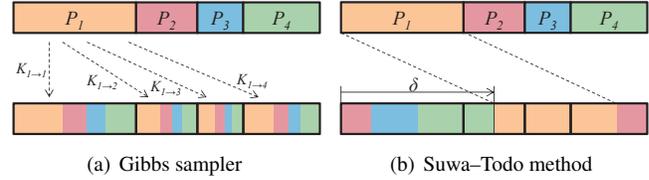


Figure 1. Example of landfill for the transition kernel of Gibbs sampler and Suwa–Todo method. Upper row depicts current distribution and lower row depicts the distribution after applying kernel. The transition kernel is visualized as moving boxes. Unlike Gibbs sampler, the kernel of Suwa–Todo method has the zero rejection rate in this example. (best viewed in color)

detailed balance is not trivial. Recently, Suwa and Todo [17] proposed a generic framework to build a non-reversible kernel without detailed balance. This method updates a single node at a time, similar to the single-site Gibbs sampler.

The single-site Gibbs sampler updates a single node from its conditional distribution $P(x_i | \mathbf{x} \setminus x_i)$. Given that we are dealing with the process of updating a single node, let us omit $(x_i | \mathbf{x} \setminus x_i)$ and denote the probability as P_a when x_i is assigned with the label a . We denote a transition kernel as $K_{a \rightarrow b}$ where a is the current label of the node i and b is the next label. The detailed balance condition is given by the following equation.

$$P_a K_{a \rightarrow b} = P_b K_{b \rightarrow a}. \quad (2)$$

A transition kernel can be visually understood as a landfill model. The kernel for Gibbs sampler is depicted in Fig. 1(a). In the landfill model, the probabilities are represented as boxes with size proportional to the probability values. These boxes move according to the transition kernel $K_{a \rightarrow b}$ while preserving the size of the probability boxes.

The design of the transition kernel is an important determinant of how fast the MCMC converges. By adopting the MH method, we can obtain a transition kernel with a more general form than that of the conditional distribution. However, the Gibbs and MH methods are limited because they design kernel while satisfying the detailed balance. Therefore, this paper aims to determine whether a better kernel can be designed and aims to identify the conditions required to achieve a better kernel.

Although defining the optimal kernel is a non-trivial task, a guideline is available in achieving a good kernel. For instance, this kernel should minimize rejection rate. The rejection rate of the Gibbs sampler can be calculated by $\sum_i K_{i \rightarrow i}$. The Suwa–Todo method proposes a non-reversible kernel that minimizes the rejection rate.

The Suwa–Todo method can be easily understood by employing the landfill model. The transition kernel for the Suwa–Todo method is illustrated in Fig. 1(b) and Alg. 1.

Algorithm 1 The transition kernel for Suwa–Todo method

- 1: Consider a node i currently labeled as a
 - 2: $\delta \leftarrow 0.5$
 - 3: **if** $a = 1$ **then** $\gamma \sim \mathcal{U}(0, C_a)$;
 - 4: **else** $\gamma \sim \mathcal{U}(C_{a-1}, C_a)$;
 - 5: $T \leftarrow \gamma + \delta$;
 - 6: **if** $T > 1$ **then** $T \leftarrow T - 1$;
 - 7: **for** $k \leftarrow 1$ to L **do**
 - 8: **if** $T < C_k$ **then** $b \leftarrow K$; break;
 - 9: **end for**
 - 10: Assign b to x_i
-

Let us consider how the transition kernel updates a single node. First, γ is randomly chosen from the uniform distribution between C_{a-1} and C_a , where $C_a = \sum_{k=1}^a P_k$. The node i is updated to the label b s.t.

$$P_{b-1} < \gamma + \delta \leq P_b \quad (3)$$

or

$$P_{b-1} < \gamma + \delta - 1 \leq P_b. \quad (4)$$

In Fig. 1(a), For example, if the current label is 1, the next label is assigned to be 2, 3, or 4 with the probability in proportion to the box sizes. If the current label is 2, the next label is assigned to be 4 or 1. If the current label is 3, the next label is always 1. If the current label is 4, the next label is 1 or 2.

The transition kernel can be written as

$$K_{a \rightarrow b} = \max(0, \min(P_a, P_b, C_a - C_{b-1} + \delta, C_b - C_{a-1} - \delta)), \quad (5)$$

In the original version of the Suwa–Todo method, δ was chosen as $\max_k P_k$. We fixed δ to be 0.5 and it is straightforward to show that the rejection rate remains the same. The rejection rate is 0 when $\max_k P_k < 0.5$ and is $(\max_k P_k - 0.5)$ otherwise. This rate is an optimum for a single-node-update MCMC. This rate is not achievable with the detailed balance condition.

The Markov chain from Suwa–Todo method converges to the target distribution by satisfying the following balance condition:

$$\sum_{j=1}^L (P_i K_{i \rightarrow j} - P_j K_{j \rightarrow i}) = 0, \quad (6)$$

which is much relaxed than the detailed balance condition is.

4. Scanline Gibbs sampler

In the single-site Gibbs sampler, only one variable is updated at a time according to its conditional distribution. This is the main reason for the slow convergence of the method. Therefore, instead of updating a single node, we update a multiple number of nodes at a time. This strategy is called the blocked Gibbs sampler.

The blocked Gibbs sampler updates a block \mathcal{B}_i from its conditional distribution $P(\mathcal{B}_i | \mathbf{x} \setminus \mathcal{B}_i)$. The union of blocks should be the same as the set of all nodes ($\bigcup_i \mathcal{B}_i = \{x_i\}$). The blocks do not need to form the disjoint set. To sample from the conditional distribution $P(\mathcal{B}_i | \mathbf{x} \setminus \mathcal{B}_i)$, a block \mathcal{B}_i , which is conditioned on other variables, should be in its specific form. The structure of the blocks is usually limited to a tree so that the calculation of marginal distribution is available in polynomial time.

There are several guidelines on how to determine the blocks. A typical trend is to choose as many nodes as possible. However, this strategy was shown to be inappropriate for grid graph optimization. We rather determine the blocks that are denoted as the disjoint set of each row. We experimentally found that the scanline blocks outperforms tree blocks, which can expand to whole images and contain about 50% of the total nodes. When a single-site update method, such as the single-site Gibbs sampler or iterated conditional modes (ICM), is applied to the grid graph, we should choose the visiting order of each node. The node maybe ordered as either a raster scan order or a random order. The raster scan order strategy outperforms the random order. Scanline blocks work in a way similar to that of the raster scan order scheme.

Let us consider a set of nodes $\mathbf{x} = \{x_1, \dots, x_m\}$ in the form of a chain. The joint distribution of \mathbf{x} can be formulated as follows:

$$P(\mathbf{x}) = \prod_{i=1}^m \phi_i(x_i) \prod_{i=1}^{m-1} \phi_{i,i+1}(x_i, x_{i+1}). \quad (7)$$

We first apply forward message passing by using the following equation:

$$\mu_{i \rightarrow i+1}(x_{i+1}) = \sum_{x_i=1}^L \phi_i(x_i) \phi_{i,i+1}(x_i, x_{i+1}) \mu_{i-1 \rightarrow i}(x_i) \quad (8)$$

The conditional distribution for a node i conditioned only on x_{i+1} can be derived as follows:

$$p(x_i | x_{i+1}) \propto \phi_i(x_i) \phi_{i,i+1}(x_i, x_{i+1}) \mu_{i-1 \rightarrow i}(x_i). \quad (9)$$

Once the joint distribution is decomposed in the form of

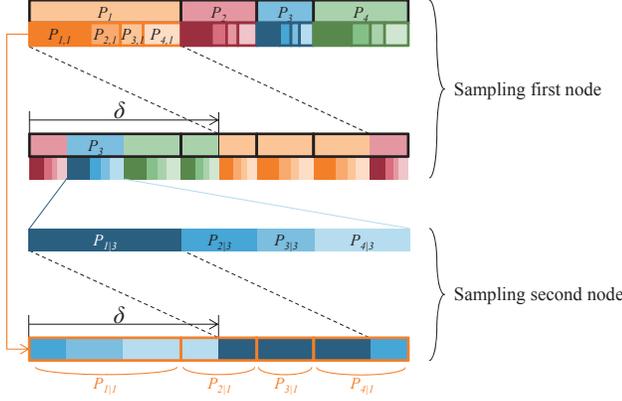


Figure 2. Example of landfill for the transition kernel of scanline sampler without detailed balance. The joint distribution is two-dimensional in this example. Upper two rows depicts the process of sampling the second variable from its marginal distribution. Lower two rows depicts the process of sampling the first variable from its conditional distribution. The transition kernel is visualized as moving boxes. (best viewed in color)

$$P(\mathbf{x}) = P(x_1|x_2)P(x_2|x_3) \cdots P(x_{m-1}|x_m)P(x_m), \quad (10)$$

exact sampling is available starting from node x_n to x_1 . Note that this form of sampling significantly differs from the single-site Gibbs sampler.

To apply exact sampling from a chain, which is chosen from grid MRF model, we performed the following process. First, a whole graph is divided into disjoint set, in which each row becomes each block. Each row is then sampled from its conditional distribution given the remaining variables. To sample a given row of nodes, we first modified the unary potential for each node to be $\tilde{\phi}_i(x_i) = \phi_i(x_i) \prod_{j \in \mathcal{N}_{out}(i)} \phi_{i,j}(x_i, x_j)$, where $\mathcal{N}_{out}(i)$ is the neighbor of the node i outside the chosen row. Aforementioned exact sampling is subsequently applied to the set of nodes.

5. Scanline sampler without detailed balance

We have introduced two ideas to accelerate the convergence of MCMC. In this section, we integrate those two ideas to propose a novel method called scanline sampler without detailed balance. In our knowledge, this is the first versatile algorithm that designs a transition kernel that can update multiple nodes without detailed balance.

The proposed method extends the Suwa-Todo method to update multiple nodes. An example is illustrated in Fig. 2. This example shows the sample from the joint distribution with two variables. The proposed method can be generalized in cases with more than two variables.

Let us consider a scanline block defined in the previous

Algorithm 2 Scanline sampler without detailed balance

- 1: Consider a set of node $\mathcal{B} = \{x_1, \dots, x_m\}$ currently labeled as (a_1, \dots, a_m)
 - 2: $\delta \leftarrow 0.5$
 - 3: Draw b_n from the marginal P_{x_m} using the Alg. 1.
 - 4: **for** $k \leftarrow (m-1)$ to 1 **do**
 - 5: Draw b_k from $P_{x_k|x_{(k+1)}}$ as following;
 - 6: **if** $a_k = 1$ **then** $\gamma \sim \mathcal{U}(0, C_{a_k|b_{(k+1)}})$;
 - 7: **else** $\gamma \sim \mathcal{U}(C_{(a_k-1)|a_{(k+1)}}, C_{a_k|b_{(k+1)}})$;
 - 8: $T \leftarrow \gamma + \delta$;
 - 9: **if** $T > 1$ **then** $T \leftarrow T - 1$;
 - 10: **for** $l \leftarrow 1$ to L **do**
 - 11: **if** $T < C_{l|b_{(k+1)}}$ **then** $b_k \leftarrow l$; **break**;
 - 12: **end for**
 - 13: Assign b_k to x_k
 - 14: **end for**
-

section. The algorithm begins with having the joint distribution $P(\mathbf{x})$ decomposed into the form of Equation (10). In the two-variable case, we can denote the joint distribution as $P(x_1, x_2) = P(x_1|x_2)P(x_2)$.

The first step is applying the Suwa-Todo method to sample x_2 from its marginal distribution $P(x_2)$. Let us say that the label of x_2 has changed from a_2 to b_2 . In the second step, we shift the probability boxes, which are in proportion with the conditional distribution $P(x_1|x_2 = a_2)$. The shifted probability boxes are divided according to the conditional distribution $P(x_1|x_2 = b_2)$. In Fig. 2, for example, if the current label is (1, 3) (blue boxes in Fig. 2), the next label of x_2 is always 1. Subsequently, x_1 is updated to be 2, 3 or 4 with proportional to divided box sizes.

The entire procedure is illustrated in Alg. 2. For brevity, we denote $P(x_1 = a_1)$ as P_{a_1} , $P(x_1 = a_1, x_2 = a_2)$ as P_{a_1, a_2} , and $P(x_2 = a_2|x_1 = a_1)$ as $P_{a_2|a_1}$. $C_{a_v|b_w}$ is defined as $\sum_{k=1}^{a_v} P_k|b_w$. Unlike in the Suwa-Todo method, in this case, obtaining the optimal δ , which minimizes the rejection rate, is non-trivial. However, setting $\delta = 0.5$ gives satisfying results in our experiments.

6. Experimental settings

6.1. Initialization

In the theoretical aspect, MCMC guarantees optimal solution without regard to initialization. However, a long convergence time is needed to obtain the optimal solution. Given that the fast annealing schedule is usually applied in practice, careful initialization of the solution is of crucial importance when employing the sampling-based optimization method. Although tailored application-specific initializations can be employed, these initializations are usually tricky and difficult to obtain. Therefore, we propose a simple and generic method to obtain good initial solutions with-

out having prior knowledge of the target application.

One of the most well-known generic initialization methods is the winner-takes-all (WTA), which was used as an initialization method for the iterated conditional modes in [18]. WTA ignores the pairwise penalty terms, and each node is assigned to the label with the lowest data cost. Since it only considers data costs, the initial solution from WTA is not satisfactory. We need better initialization method.

When considering a scanline block updating scheme, the natural extension of WTA would be from a single node to a row of nodes. The proposed winner-takes-scanline (WTS) algorithm ignores the pairwise penalty terms $\phi_{i,j}(x_i, x_j)$ if nodes i or j is not in the row. The nodes in the row are assigned to labels with the lowest cost. The unary terms $\phi_i(x_i)$ and the pairwise terms $\phi_{i,j}(x_i, x_j)$ are considered only if nodes i and j are both in the row. Optimal labeling can be achieved by dynamic programming in polynomial time. In experimental section, we include initialization time to the total running time.

6.2. Annealing schedule

Our main objective is not the sampling but rather maximum a posterior (MAP) estimation. To achieve MAP solution with sampling method, we need to apply simulated annealing (SA) scheme together with the proposed sampler. In SA, the target distribution is changing for each iteration. The target distribution at iteration k is given by

$$P_k(\mathbf{x}) = \frac{1}{Z_k} \{P(\mathbf{x})\}^{\frac{1}{T_k}}, \quad (11)$$

where Z_k is a normalizing constant and T_k is a temperature. The temperature T_k at each iteration is determined by

$$T_k = T_1 c^{k-1}, \quad (12)$$

where $c < 1$.

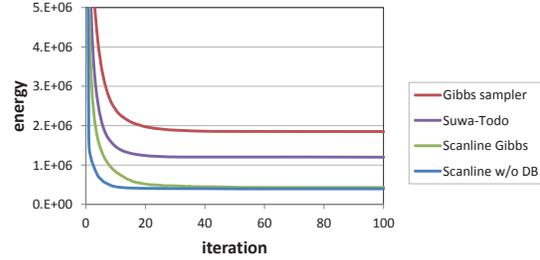
The choice of initial temperature T_1 and the decaying factor c is important to achieve good solutions. There is no method for finding a suitable annealing schedule for a whole range of problems. Therefore, we consider following process to determine T_1 and c .

We first initialize the solution using the aforementioned scheme. Having the initial solution, the average of pixel-wise conditional log-probability are examined as following.

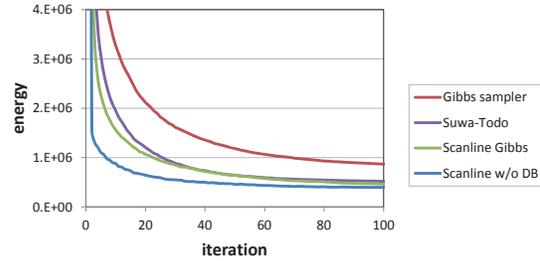
$$q = -\frac{1}{N} \sum_{i=1}^N \log \hat{P}(x_i | \mathbf{x} \setminus x_i), \quad (13)$$

where $\hat{P}(x_i | \mathbf{x} \setminus x_i)$ is unnormalized probability calculated from unary and pairwise potential. The initial temperature is determined as

$$T_1 = q\alpha, \quad (14)$$



(a) $\alpha = 0.1$ and $c = 0.9$ (Low temperature and fast decay)



(b) $\alpha = 0.1$ and $c = 0.99$ (Low temperature and slow decay)

Figure 3. Comparison results of the convergence speed of each algorithm.

where α is a value in $[0.5, 2]$. α is experimentally chosen for each dataset.

We terminate the algorithm if the probability of the solution does not increase over 200 iterations. Maximum iteration is fixed to 2000.

7. Experiments

7.1. Convergence speed analysis

The main objective of this research is to increase the convergence speed of Markov chain. To analyze the performance improvement, we evaluated different sampling algorithms. They include Gibbs sampler, Suwa-*Todo* method, scanline Gibbs sampler, and scanline sampler without detailed balance.

There are two strategies for accelerating MCMC. Suwa-*Todo* method improve convergence rate by breaking detailed balance. Scanline Gibbs sampler updates multiple nodes at a time. And scanline sampler without detailed balance exploits two advantages simultaneously.

For evaluation, we tested various methods on the stereo matching problem. As an input, ‘Tsukuba’ images are used. And we used energy model from the middlebury benchmark [18].

To effectively observe the differences in convergence speed, we did not use WTS scheme for the initialization because WTS provides good initial solution with low en-

ergy. Instead, we initialize the solution with random numbers. For different algorithms, same initial solution was used. Also, to maximize the difference, we start from low initial temperature ($\alpha = 0.1$).

We designed two different situation, $c = 0.9$ and $c = 0.99$. The comparison results are illustrated in Fig. 3. The energy decrease of each algorithm during the first 100 iteration is plotted. The running time for a single iteration is approximately same for each method. So we choose to present plots with iteration to exclude the effect from implemnetal errors.

Fig. 3(a) shows the results of annealing schedule with $\alpha = 0.1$ and $c = 0.9$. The Markov chain starts from low temperature and decrease fast. Since conventional Gibbs sampler has slower than others, it ends up with being captured in local minima. Suwa-Todo method works better than Gibbs sampler. It achieved lower state but it also ends up with bad local minima. Two scanline-based samplers achieved similarly good results. However, it shows that scanline sampler without detailed balance decreases energy faster than scanline Gibbs sampler does.

Fig. 3(b) shows the results of annealing schedule with $\alpha = 0.1$ and $c = 0.99$, where the decay of the temperature are slower than previous situation. In this figure, no algorithm seems to be captured in bad local minima with high energy costs. However, they are significantly different in terms of convergence speed. For example, scanline sampler without detailed balance obtains the energy lower than $(10)^6$ after 7 iterations, while Gibbs sampler obtains it after 69 iterations.

7.2. OpenGM2 benchmark

We evaluated the proposed method on OpenGM2 benchmark [8]. This benchmark includes various types of energy models for various applications. Among them, we applied our algorithm on 4-neighborhood grid graph models. Specifically, the experiments were conducted on color segmentation (*color-seg-n4*, 9 instances), object segmentation (*object-seg*, 5 instances), and inpainting (*inpainting-n4*, 2 instances). For the detailed explanation of the energy models, please refer to [8]. All the experiments were performed on the Intel i5-2500 3.3GHz CPU and 8 GB RAM.

We evaluated four MCMC based method: Gibbs sampler, Suwa-Todo method, scanline Gibbs sampler, and scanline sampler without detailed balance. For further comparison, we also report the results of other methods including FastPD [15], belief propagation (LBP/BPS) [18], α -expansion, $\alpha\beta$ -swap [6, 14], TRW [13], bundle methods (BUNDLE-A, BUNDLE-H) [9], multicut solver (MCA) [10], TRBP [21], and lazy flipper (-LF2) [2]. Their results are directly borrowed from [8].

The final results are summarized in Tab. 1–3. They contain the running time and final energy which are averaged

Table 1. inpainting-n4

Algorithm	mean run time	mean value
Scanline w/o DB	18.01 sec	454.75
Scanline Gibbs	17.82 sec	454.75
Suwa-Todo	9.45 sec	454.75
Gibbs sampler	8.90 sec	459.07
FastPD	0.02 sec	454.75
FastPD-LF2	0.37 sec	454.75
mrf-LBP-LF2	5.33 sec	475.56
mrf-BPS	2.13 sec	454.35
mrf-EXPANSION	0.02 sec	454.35
mrf-SWAP	0.02 sec	454.75
mrf-TRWS	2.19 sec	490.48
ogm-BUNDLE-A	76.84 sec	455.25
ogm-BUNDLE-H	36.55 sec	455.25
ogm-SUBGRAD-A	47.20 sec	455.25
ogm-ILP	1816.11 sec	454.75
ogm-LBP	76.21 sec	480.27
MCA	1810.59 sec	4618.38
MCA (6h)	12262.31 sec	474.38
ogm-TRBP	90.46 sec	480.27
TRWS-LF2	4.68 sec	489.3

for each dataset.

Over all experiments, scanline sampler without detailed balance outperforms all other MCMC methods. Suwa-Todo method always outperforms Gibbs sampler. It certainly reveals the advantage of using non-reversible kernels. Scanline Gibbs sampler typically achieves better results than both single-site Gibbs sampler and Suwa-Todo method.

Inpainting ‘*inpainting-n4*’ dataset contains two instances. Tab. 1 shows the results for ‘*inpainting-n4*’ dataset. The α value for the initial temperature was set to 0.5. c is fixed to 0.995 for every experiment in this section. All the method except Gibbs sampler was able to find the good solutions. All three methods converged in a first few iterations, but they spend longer time only to wait for the stopping condition.

Object segmentation ‘*object-seg*’ dataset contains five instances. Tab. 2 shows the results for ‘*object-seg*’ dataset. The α value for the initial temperature was set to 2. There is significant improvement over all other sampling-based method.

Color segmentation ‘*color-seg-n4*’ dataset contains nine instances. Tab. 3 shows the results for ‘*color-seg-n4*’ dataset. The α value for the initial temperature was set to 0.5. Our method achieved better results than other sampling methods also in this dataset

Table 2. object-seg

Algorithm	mean run time	mean value
Scanline w/o DB	51.05 sec	32243.75
Scanline Gibbs	47.62 sec	32318.37
Suwa-Todo	77.31 sec	35623.84
Gibbs sampler	72.35 sec	35607.25
FastPD	0.17 sec	31317.60
FastPD-LF2	2.88 sec	31317.60
mrf-LBP-LF2	41.26 sec	32400.01
mrf-BPS	15.72 sec	35775.27
mrf-EXPANSION	0.43 sec	31317.23
mrf-SWAP	0.34 sec	31323.18
mrf-TRWS	16.21 sec	31317.23
ogm-BUNDLE-A	215.68 sec	31317.31
ogm-BUNDLE-H	346.56 sec	31317.23
ogm-SUBGRAD-A	365.16 sec	31424.55
ogm-ILP	788.71 sec	33884.39
ogm-LBP	445.85 sec	32663.86
MCA	278.51 sec	31317.23
ogm-TRBP	856.63 sec	32663.86
TRWS-LF2	28.43 sec	31317.23

Table 3. color-seg-n4

Algorithm	mean run time	mean value
Scanline w/o DB	80.62 sec	20044.44
Scanline Gibbs	67.95 sec	20050.93
Suwa-Todo	127.71 sec	20360.51
Gibbs sampler	127.27 sec	20368.35
FastPD	0.35 sec	20034.80
FastPD-LF2	13.61 sec	20033.21
mrf-LBP-LF2	63.82 sec	20053.25
mrf-BPS	32.92 sec	20094.03
mrf-EXPANSION	1.24 sec	20031.81
mrf-SWAP	0.86 sec	20049.90
mrf-TRWS	33.15 sec	20012.18
ogm-BUNDLE-A	692.39 sec	20024.78
ogm-BUNDLE-H	1212.24 sec	20012.44
ogm-SUBGRAD-A	1179.62 sec	20027.98
ogm-LBP	1887.89 sec	20054.26
MCA	982.36 sec	20527.37
MCA-6h	1244.30 sec	20012.14
ogm-TRBP	2516.54 sec	20054.06
TRWS-LF2	89.83 sec	20012.17

8. Conclusions

Markov chain Monte Carlo (MCMC) has been widely used for the variety of areas. It has also been used for maximum a posteriori problems on Markov random field model. Despite of its generality and elegancy, it has often been regarded as less powerful than deterministic algorithms. The main reason is its slow convergence. In this paper, we introduced two important ideas which significantly improve the convergence speed of Markov chain. Those ideas are breaking detailed balance and updating multiple nodes at a time.

Although detailed balance condition provides us great simplicity in designing a kernel, it is not a necessary condition for MCMC. By breaking detailed balance, more efficient kernel is to be available. Second idea of updating multiple nodes has been pursued in the computer vision literature but they are often restricted. We proposed most generic framework for updating multiple nodes. Finally, we integrated those two ideas to build a new efficient algorithm called scanline sampler without detailed balance. We evaluated different version of MCMC methods and showed our proposed method outperform other sampling based algorithms.

Acknowledgments

This research was supported in part by the Advanced Device Team, DMC R&D, Samsung Electronics.

References

- [1] K. Alahari, P. Kohli, and P. Torr. Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In *CVPR*. IEEE, 2008.
- [2] B. Andres, J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. The lazy flipper: Efficient depth-limited exhaustive search in discrete graphical models. In *ECCV*, pages 154–166. Springer, 2012.
- [3] A. Barbu and S. C. Zhu. Multigrid and multi-level Swendsen-Wang cuts for hierarchic graph partition. In *CVPR*, 2004.
- [4] A. Barbu and S. C. Zhu. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *PAMI*, 27(8):1239–1253, 2005.
- [5] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26(9):1124–1137, 2004.
- [6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
- [7] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *PAMI*, 20(6):721–741, 1984.
- [8] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, et al. A comparative study of modern inference techniques for discrete energy minimization problems. In *CVPR*, 2013.
- [9] J. H. Kappes, B. Savchynskyy, and C. Schnörr. A bundle approach to efficient map-inference by lagrangian relaxation. In *CVPR*, pages 1688–1695. IEEE, 2012.
- [10] J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schn. Globally optimal image partitioning by multicuts. In *EMM in CVPR*, pages 31–44. Springer, 2011.
- [11] W. Kim, J. Park, and K. M. Lee. Stereo matching using population-based MCMC. *IJCV*, 83(2):195–209, 2008.
- [12] P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV*, 2005.

- [13] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.
- [14] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.
- [15] N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *PAMI*, 29(8):1436–1453, 2007.
- [16] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
- [17] H. Suwa and S. Todo. General construction of irreversible kernel in markov chain monte carlo. *arXiv preprint arXiv:1207.0258*, 2012.
- [18] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *PAMI*, 30(6):1068–1080, 2008.
- [19] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *ICCV*, 2003.
- [20] K. S. Turitsyn, M. Chertkov, and M. Vucelja. Irreversible monte carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4):410–414, 2011.
- [21] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.